FEDERAL AVIATION ADMINISTRATION
AAR-100 (Room 907)
800 Independence Avenue, S.W.
Washington, D.C. 20591

Tel: 202-267-8758
Fax: 202-267-5797
william.krebs@faa.gov

January 10th, 2003

From: General Aviation Human Factors Program Manager, AAR-100
To: General Aviation TCRG

Subj: GENERAL AVIATION HUMAN FACTORS FIRST QUARTER '03 REPORT

1) Below is a summary of the projects that address the general aviation TCRG requirements.

2) FY03 projects:

   a) Causal factors of accidents and incidents attributed to human error.

   The existing GA human error database was updated to include calendar year 1999 and several earlier accidents that had been finalized since the last HFACS update nearly a year ago. Following the recent update, a report was completed entitled, "Reshaping the way we look at General Aviation Accidents using the Human Factors Analysis and Classification System." Currently, the manuscript is undergoing review by AFS-801 and AAM-1 prior to presentation at the International Symposium of Aviation Psychology to be held in Dayton, OH in April. The paper details our analyses of 14,571 GA accidents that occurred between 1990-99. Overall, skill-based errors (primarily technique errors) were associated with nearly four out of every five GA accidents since 1990, followed by decision errors (37%) and perceptual errors (less than 10%) regardless of whether the accident resulted in a fatality. On the other hand, violations of the rules were more common among fatal (32%) than non-fatal (10%) accidents. Finally, there was little difference between FAA geographic regions in the types of unsafe acts committed by GA pilots involved in accidents. In particular, there appeared to be no differences in the Alaska region as compared to the rest of the U.S. when using HFACS. These analyses provide unique insight into the genesis of GA accidents. Implications for GA initial and recurrent training were discussed (Please see the attached manuscript).

   In addition to the overall GA analysis, a separate human error analysis comparing homebuilt accidents (n=1976) with all other GA accidents (16,183) was conducted for AFS-300. While maintenance error involvement was higher for homebuilt (313 accidents, 15.8%) than all other GA accidents (1199, 7.4%), the distribution of unsafe acts was noticeably lower for homebuilt accidents (Table 1).

Table 1. Number and percentage of homebuilt and all other GA accidents associated with at least one instance of a given unsafe act.

Unsafe Acts	*All other GA*		**Homebuilt**	
Errors	Frequency	Percentage	Frequency	Percentage
Decision errors	4,692	30.7	430	21.8
Skill-based errors	10,243	63.3	1,168	59.1
Perceptual Errors	1,106	6.8	80	4.0
Violations	1,998	12.3	166	8.4

It should be noted that these findings are preliminary, and while they suggest that the percentage of specific errors and violations associated with homebuilt aircraft are lower than other GA accidents, the proportion of fatalities is considerably higher (30.2% of the homebuilt accidents examined resulted in a fatality compared with 18.8% of other GA accidents). A final report is planned in the next quarter.

- Dr. Scott Shappell (CAMI) and Dr. Douglas Wiegmann (University of Illinois) were presented the Flight Safety Foundation's Admiral Luis de Florenz Flight Safety Award for their for the development and application of the Human Factors Analysis and Classification System (HFACS) as a reliable taxonomy for investigating human factors data in aviation accidents and incidents, revealing previously unknown human-error trends.

All indications indicate that this project is on track to complete the milestones as planned.

b) Comparison of the Effectiveness of a Personal Computer Aviation Training Device, a Flight Training Device and an Airplane in Conducting Instrument Proficiency Checks.

During this quarter 66 pilots had started the study, and 292 sessions had been scheduled. A total of 43 pilots had completed Instrument Proficiency Check #1 and 38 pilots had completed Instrument Proficiency Check #2 (thus 38 pilots had completed their participation in the study with 5 more requiring one more IPC to complete their participation). Additionally, participants as part of the requirements to complete the study had completed the following sessions:

Familiarization sessions – 50 in the aircraft, 51 in the PC-based Aviation Training Device, 51 in the Flight Training Device.

Training sessions – 1 in the aircraft, 19 in the PC-based Flight Training Device, 39 in the Flight Training Device.

All indications indicate that this project is on track, however this project does not have an execution plan and it is unclear how the project's deliverable will meet the sponsor's objectives.

Recommend that this project be evaluated at the TCRG to determine who is the sponsor point of contact and how this project will meet the TCRG objectives.

c) Credit for Instrument Rating in a Flight Training Device or Personal Computer

   i. Phase I: Survey UAA, Part 61, and Part 141 institutions. Report submitted to AAR-100 on December 31st, 2002.

      Project completed.

   ii. Phase II: Capabilities of FTDs/PCATDs. Researchers identified and targeted 149 training organizations, 65 universities, 41 Part 141 schools, and 43 Part 61 schools. The list was based on the Phase I study survey findings. The researchers distributed the survey on December 16th and anticipate the surveys to be returned by the end of January 2003. The survey responses will be analyzed in February with the final report delivered in March 2003.

      Indications are that there are minor risks to the activity being completed as planned. The deliverable will be delayed to March 2003, a slip of two months.

   iii. Phase III: Transfer of Training Effectiveness of a Flight Training Device (FTD). Thirty-eight students completed the AVI 130 Basic Instrument course and took the final check ride for the course. A total of 24 students passed the check ride on the first attempt whereas 13 required a second attempt. One student failed the check ride on the second attempt and 3 others failed to complete the course; all were recommended for a remedial course, AVI 102.

Table 1. Summary of outcomes for the 6 groups to date (Fall Semester, 2002)

	Airplane Only	PCATD 5 Hours	FTD 5 Hours	FTD 10 Hours	FTD 15 Hours	FTD 20 Hours
Number of Students	**7**	**6**	**6**	**6**	**7**	**6**
Number passing check ride on first attempt	3 *(42.8%)*	5 *(83.3%)*	4 *(66.7%)*	4 *(66.7%)*	6 *(85.7%)*	2 *(33.3%)*
Number passing check ride on second attempt	4 *(57.2%)*	1 *(16.7%)*	2 *(33.3%)*	2 *(33.3%)*	0	4 *(66.7%)*
Students failing both check rides *(or *failing to complete course)*	0	0	*1**	*1**	1 (14.3%)	*1**
Mean dual-instruction hours required to complete course for those passing check rides	20.74 *(N = 7)*	18.70 *(N = 6)*	18.37 *(N = 6)*	18.85 *(N = 6)*	19.88 *(N = 6)*	17.58 *(N = 6)*
Variance associated with mean completion hours	7.90	3.06	6.90	12.80	3.03	11.58

Note: This lesson is the final check ride for AVI 130.

This quarter, forty-one AVI 130 Basic Instruments students started the project this semester; 37 students successfully completed the semester course. A survey to collect opinions concerning the effectiveness of the cross-country scenarios was completed by the flight instructors. The Institute received notification that the exemption from the flight hours of Part 141 has been approved.

Indications are that this activity is on track.

d) <u>Developing And Validating Criteria for Constraining False & Nuisance Alerts For Cockpit Display Of Traffic Information Avionics</u>.

The literature search has revealed a number of studies, which had evaluated—in one form or another—aircraft conflict alert logic in the context of either a CDTI for pilots or a traffic display for controllers. These studies (down the rows) are summarized in the attached table, which emphasizes a number of features (across the columns) of each study. A review of the studies in the table indicates that all algorithms received some form of validation. Most importantly, some of the validations have examined specifically the impact of changing parameters or algorithm attributes on false alert rate. For example, it appears

- that increasing the size of the protected zone (PZ) and increasing the look-ahead time (LAT) will increase the false alarm rate (FAR),
- that educing FAR by adjusting the criterion for alert must, by definition, increase the miss rate (MR) or delay the initiation of alert, and
- that interactions between these parameters are not always intuitive.

However, only three particular algorithms appear to have been validated by PIL simulations (and therefore contain human performance data). These are the Yang and Kuchar algorithm (3 studies), the NLR algorithm (Hoekstra, 1 study), and the Illinois algorithm (several studies, of which Wickens, Gempler and Morphew, 2000, appears to be most relevant). Of these algorithms, all have included multi-level alerts (although none have manipulated the number of levels), but only the first two have included explicit assumptions within the algorithm, regarding uncertainty growth.

Most critically, while all studies have presumably examined pilot response, with the exception of Wickens et al. (2000), none of the studies reported the response of pilots to algorithm errors, such as false alarms, nor how those error actually are found to impact pilot performance as their frequencies are increased (i.e., by increasing the diameter of the protected zone, or by increasing the look-ahead time). Wickens et al. (2000) examined algorithm errors in terms of misses, rather than false alerts, and reported that while there was a mild disruption to conflict resolution performance (presumably reflecting a delay in conflict detection, although this was not explicitly measured), pilots generally adapted well to the imperfections of predicting the uncertain future, and their explicit level of trust (estimated system reliability) was closely calibrated to the actual level of reliability.

Our continued literature search has provided scant results in terms of relevance to the hypotheses set for this research. Only one paper relevant to (in support of) hypothesis 1 (operators' tolerance for false alarms can be increased by improving their general awareness of the traffic situation) has been found, by Cotté, Meyer, and Coughlin (2001), and only four papers relevant to (in support of) hypothesis 2 (false alarm tolerance can be improved by increasing the resolution of) by Gupta, Bisantz, and Singh (2001), John and Manes (2002), Sorkin, Kantowitz, and Kantowitz (1988), and Masalonis and Parasuraman (2000). Of these, only two (Sorkin et al. and Masalonis and Parasuraman) were relevant to the aviation domain. Wickens, Gempler, & Morphew's (2000) experiment 1 supports hypothesis 1, but is only partially consistent with hypothesis 3 (operators' performance can be significantly improved by displaying probabilistic information to them in a form that is easy to perceive and understand and that can be readily used in their tasks). What is noteworthy is the lack of empirical data in areas that are most critical to human factors certification of new technologies.

We conclude from this review that there is a strong need for PIL studies that (1) vary algorithm and geometry properties across a range of plausible values; (2) measure pilots' both subjective (i.e., trust) and objective (e.g., detection latency) responses, to both misses and false alerts (that inevitably must occur and trade-off), for LATs of sufficient magnitude to allow the pilot to chose a safe avoidance maneuver, and (3) examine empirically how pilot response is or can be mitigated by multi-stage alerts, by conflict status displays, and by algorithm training (improved mental models and situation awareness).

Work on this project is progressing in four distinct and parallel areas:

1. We are developing an experimental, PIL, simulation setup to study unaided pilots' performance in conflict detection; the results will guide us in development of a cognitive model, which in turn will further our understanding of pilot performance with respect to their ability to make accurate judgments on collision risk based on CDTI information *sans* automated alerts. This will be critical in assessing the pilots' performance in response to alerts, both false and real, and allow for prediction of mistrust caused by too many false alarms as well as prediction of complacency resulting from highly reliable systems or very low conflict base rates. This experiment must therefore be considered as a prerequisite to the development of any human factors certification criteria for CASs.

2. We are developing methods to evaluate (e.g., by numerical simulations) the available conflict detection, alerting, and resolution (CDA&R) algorithms. I particular, we are identifying the *dependent* variables, or outcome variables, of the algorithms that are most relevant to human performance (e.g., trust, situation awareness, and maneuvering performance). Further work in this area will involve development of criteria for these variables.

3. We are continuing the literature review on human factors certification in general and development of a framework for human factor certification of CDTIs.

4. We are expanding the CDTI summary table (see attached) to include also ground-based (i.e., ATC) CDA&R algorithms. This work will allow for examination of the congruence of airborne and ground-based systems and consequences of incompatibilities between these.

Indications are that this activity is on track.

e) Low Visibility and Visual Detection. Grant submitted and under review. Awaiting passage of continual resolution to fund grant.

William K. Krebs